

PERFORMANCE COMPARISON BETWEEN STEREAUSIS AND INCOHERENT WIDEBAND MUSIC FOR LOCALIZATION OF GROUND VEHICLES

September 1999

Tien Pham
U.S. Army Research Laboratory
Adelphi, MD 20783-1197

Shihab Shamma and Phil Brown
University of Maryland
College Park, MD 20742

ABSTRACT

In this paper, we present experimental results comparing the *incoherent wideband MUSIC* (IWM) algorithm developed by the Army Research Laboratory (ARL)^{1, 2} and the *stereausis* algorithm developed by the University of Maryland (UMD)³ for the purpose of performing acoustic direction-of-arrival (DOA) estimation of ground vehicles. We discuss the motivating factors behind the use of auditory-inspired techniques such as stereausis for performing localization, namely, robustness and low complexity. Robustness is important because the acoustic signatures of the ground vehicles can vary significantly under different environmental conditions. We know that a human, with only two ears (sensors), can perform source separation and localization extremely well in complex environments (e.g., the cocktail party effect). Low complexity is important as well, because the algorithm will be used in real-time, unattended acoustic ground sensor applications. With the use of recently developed *aVLSI* cochlear chips,⁴ outputs from 128 auditory filter channels can be used for performing the stereausis algorithm in real time. For comparison, we will use IWM as the baseline and compare the DOA results of stereausis to that of IWM. We show raw DOA results with respect to the GPS truth data of the ground vehicles and discuss issues such as accuracy, robustness with respect to noise, number of sensor elements, computational complexity, and algorithm implementation.

1. INTRODUCTION

The Acoustic Signal Processing Branch at the Army Research Laboratory (ARL) is working with the Neural Systems Laboratory (NSL) at the University of Maryland (UMD) on applying auditory-inspired signal processing techniques to battlefield acoustic problems. In particular, we are interested in binaural processing and how it helps humans analyze complex sounds in an environment that includes multiple sound sources, multiple

¹ T. Pham and B. Sadler, "Adaptive wideband aeroacoustic wideband array," *8th IEEE SP Workshop on Statistical Signal and Array Processing*, pp. 295-298, June 1996.

² T. Pham and B. Sadler, "Wideband acoustic array processing to detect and track ground vehicles," *1st Annual ARL Sensors and Electron Devices Symposium*, pp. 151-154, January 1997.

³ S. Shamma and *et al*, "Stereausis: Binaural processing without neural delays," *Journal of Acoustical Society of America*, Vol. 83, No. 3, pp. 989-1006, 1989.

⁴ M. Erturk and S. Shamma, "A neuromorphic approach to the analysis of monaural and binaural auditory signal," *2nd European Workshop in Neuromorphic Systems*, Scotland, September 1999.

Form SF298 Citation Data

Report Date <i>("DD MON YYYY")</i> 00091999	Report Type N/A	Dates Covered (from... to) <i>("DD MON YYYY")</i>
Title and Subtitle Performance Comparison Between Stereausis and Incoherent Wideband Music for Localization of Ground Vehicles		Contract or Grant Number
		Program Element Number
Authors Pham, Tien; Shamma, Shihab; Brown, Phil		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) U.S. Army Research Laboratory Adelphi, MD 20783-1197 University of Maryland College Park, MD 20742		Performing Organization Number(s)
Sponsoring/Monitoring Agency Name(s) and Address(es)		Monitoring Agency Acronym
		Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes		
Abstract		
Subject Terms		
Document Classification unclassified		Classification of SF298 unclassified
Classification of Abstract unclassified		Limitation of Abstract unlimited
Number of Pages 12		

echoes, and moving sources. Using basically two sensors (ears) with less than one foot separation, the human can perform localization using binaural and monaural cues from interaural time differences (ITD), interaural level differences (ILD), spectral notches created by the pinna, and head movements. ITD and LTD provide azimuth information for low frequencies and high frequencies, respectively; pinna cues provides elevation and front/back information; and head movements provide front/back information. We are interested in implementing and comparing the stereausis algorithm developed at UMD with the high-resolution direction finding algorithms developed at ARL. Stereausis is an auditory-inspired processing technique based on the same fundamentals as stereopsis in vision; the main advantage of stereausis over other binaural processing techniques, such as cross-correlation, is computational complexity.³ With the use of recently developed *aVLSI* cochlear chips,⁴ outputs from 128 auditory filter channels can be used for performing the stereausis algorithm in real time in ARL's acoustic sensor testbed.⁵ In this paper, we describe the current work at ARL and UMD in acoustic wideband array processing for direction finding and tracking of ground vehicles using small baseline arrays.^{6, 7} We present simulation and experimental results comparing quantitatively and qualitatively incoherent wideband MUSIC (IWM)^{1, 2} and stereausis for different signal-to-noise ratios (SNRs).

2. ALGORITHM FORMULATION AND IMPLEMENTATION

2.1. INCOHERENT WIDEBAND MUSIC

A natural extension of the narrowband signal subspace algorithm is to combine narrowband beampatterns over many temporal frequencies.⁸ This approach is useful for acoustic signatures of ground vehicles because, there are sufficient SNRs in multiple frequency components (i.e., engine harmonics) so that a narrowband method such as MUSIC yields good results independently for each frequency. IWM is just the wideband extension of the narrowband MUSIC algorithm over a set of peak frequencies and the specific algorithm formulation and implementation can be found in papers by Pham and Sadler.^{1, 2} In general, the wideband approach provides processing gains in terms of accuracy and beampattern sharpness over narrowband processing. Since most vehicles of interest have diesel engines, they exhibit pronounced harmonic structures corresponding to the number of cylinders and the engine firing rates. The harmonic structure can be modeled as a sum of high SNR narrowband frequency components, existing for the most part between 20 and 250 Hz. Thus, given adequate SNR, IWM performs well and produces sharp and distinct peaks in the beampattern. However, the incoherent approach is not statistically stable because, low SNRs, multipath, and poor frequency selection can degrade IWM's performance significantly. For example, inclusion of low SNR frequency bins from noise tends to degrade the overall sharpness and introduce spurious peaks in the beampattern.^{1, 2}

⁵ N. Srour and J. Robertson, "Remote netted acoustic detection system: Final report," (U) ARL-TR-607, U.S. Army Research Laboratory, May 1995.

⁶ S. Shamma, D. Depireux, and P. Brown, "Signal processing in battlefield acoustic sensor arrays," *1998 Meeting of The IRIS Specialty Group on Acoustic and Seismic Sensing*, APL/JHU, September 1998.

⁷ S. Shamma *et al*, "Signal Processing in Battlefield Acoustic Sensor Arrays," *3rd Annual ARL Sensors and Electron Devices Symposium*, pp. 99-105, February 1999.

⁸ G. Su and M. Morf, "The signal subspace approach for multiple wide-band emitter location," *IEEE Trans. ASSAP*, Vol. 31, No. 6, pp. 1502-1522, December 1983.

2.2. STEREAUSIS

UMD has proposed to use stereausis to perform direction-of-arrival (DOA) estimation because it has relatively low complexity compared to other binaural processing methods, such as the cross-correlation's based methods on Jeffress's coincident detector network.⁹ The fundamental difference between stereausis and the more common binaural processing schemes is the use of spatial correlations instead of temporal correlations to extract the binaural cues. The direct implication of this fundamental difference is that stereausis does not require neural delays. The fast computation from stereausis is due to the absence of neural delays (i.e., no computation of cross-correlations at different delays). The stereausis network combines the ipsilateral (near) input and contralateral (far) input by a simple ordered matrix of operations (see figure 1 (a)). In other words, the activity of node i from the ipsilateral input (i.e., the output of the i^{th} channel from a bank of cochlear filters for the near sensor x) is compared to the activity of node j from the contralateral input (i.e., the output of the j^{th} channel from a bank of cochlear filters for the far sensor y). The output is defined as $O_{ij} = C(x_i, y_j)$, where $C(\cdot)$ is a correlation measure and can take on many forms.^{5, 6}

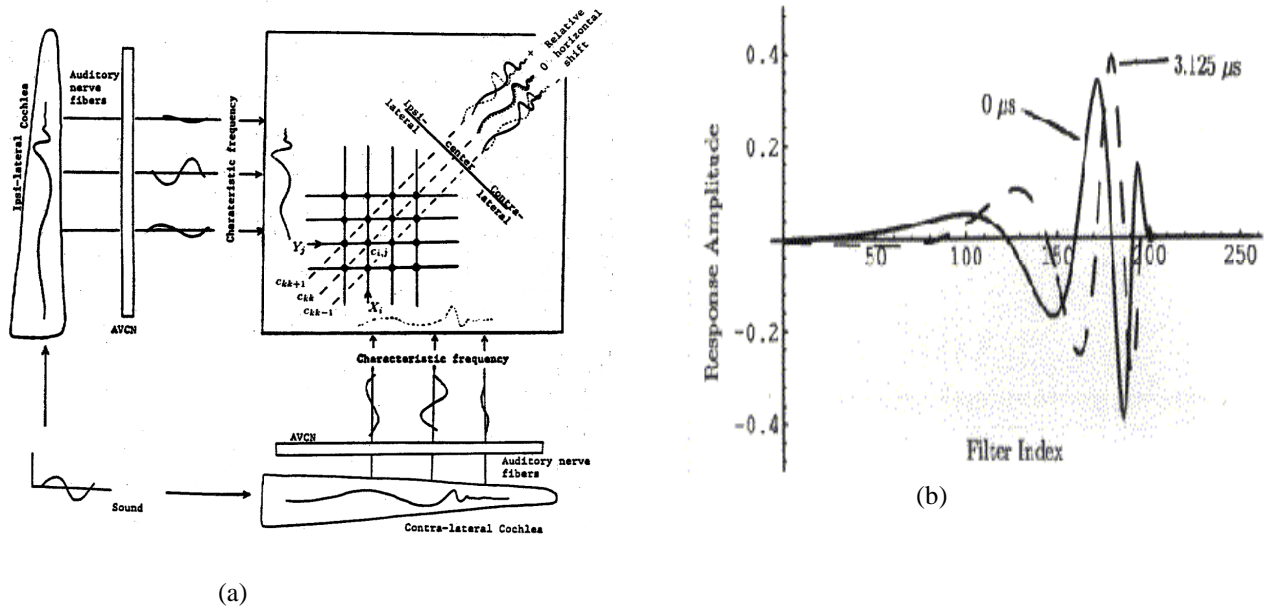


Figure 1. (a) Schematic of the stereausis-processing network showing how the ipsilateral input are correlated with the contralateral input. (b) An example of a traveling wave (ipsilateral (solid line) and contralateral (dashed line)) along the basilar membrane for a binaurally delayed tone

It is known that binaural processing at low frequencies primarily depends on ITDs and processing at higher frequencies depends on ILDs. The stereausis network shown in figure 1 (a) can process both ITD and ILD cues by using different correlation measures $C(\cdot)$.³ However, for the ground vehicle problem, the frequency range of interest is $[20, 250]$ Hz, which is low; therefore, only one correlation measure is needed for ITDs to perform DOA estimation. In fact, for the analysis shown below, the correlation measure is simply $C(x_i, y_j) = x_i y_j$. Figure 1 (b) shows a schematic of a traveling wave due to a single tone in the ipsilateral cochlea (solid line) and contralateral

⁹ L. Jeffress, "A place theory of sound localization," *J. Comp. Physio. Psych.*, Vol. 61, pp. 468-486, 1948.

cochlea (dashed line). The tone is binaurally delayed, so the sound wave propagates along the basilar membrane at two different phases, corresponding to two different time delays. DOA estimates can be derived from phase (time) delays in the disparity axes. An example of how stereausis works is illustrated in figure 2 for a signal consisting of three tones. When there is no binaural delay, the output of the stereausis network should look like the top left plot, with all three tones lining up on the main diagonal. When a delay of 7 ms is introduced between the two inputs, the tones shift away from the main diagonal. The high-frequency tones (near the center of the top right plot) have the most phase shifts for a given delay, as expected (see lower right plot). The actual ITD is calculated from the disparity plots shown in the two lower plots. Specific details of stereausis with algorithm formulation and implementation can be found in papers by Shamma *et al.*^{3, 5, 6}

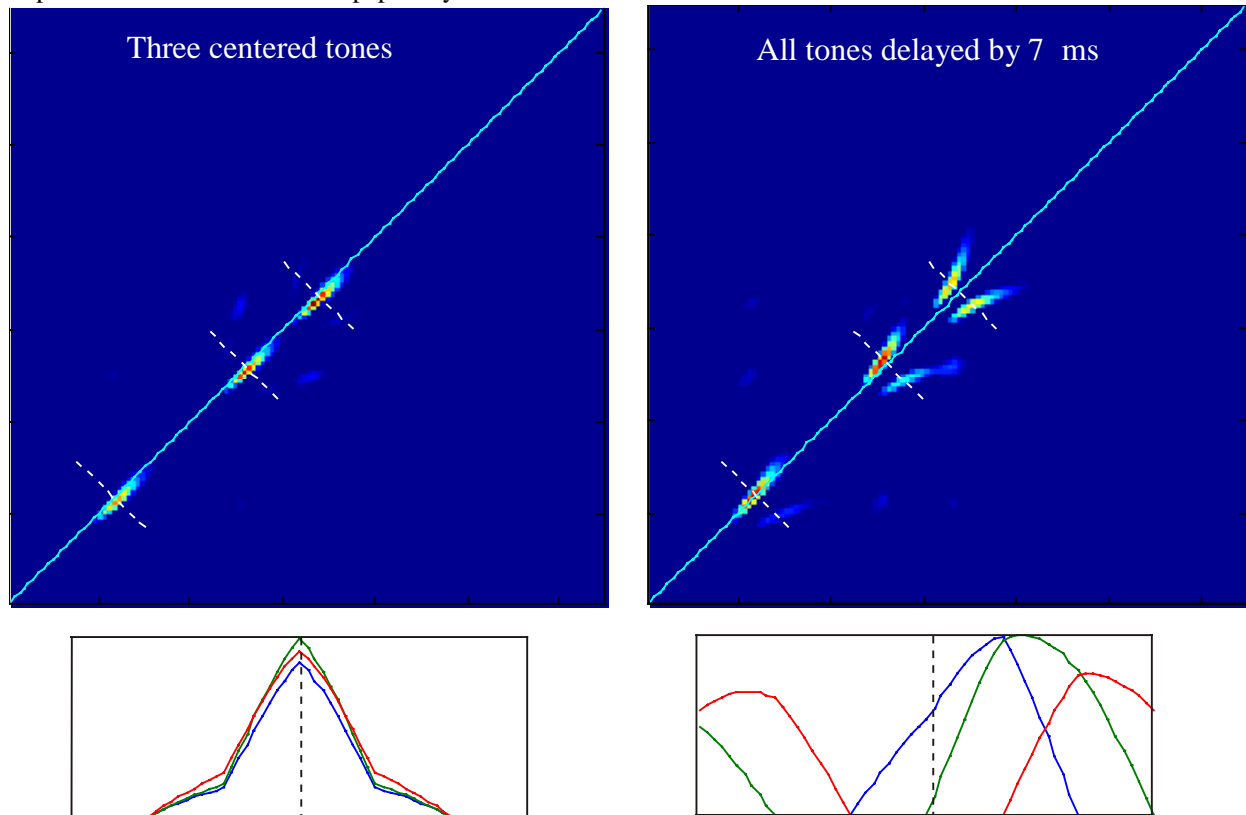


Figure 2. Stereausis output plots for three tones with no binaural delay (left plot) and with a delay of 7 ms (right plot). The bottom figures show the phase (time) delays across the three disparity axes (dashed lines) corresponding to the three tones.

3. ANALYSIS

Acoustic signature of moving tanks from a seven-element circular sensor array (6 microphones equally spaced around a circle of radius 4 ft, and one microphone at the center of the array) are used for algorithm performance evaluation. Figure 3 shows a comparison for 5 s of data between the typical spectrogram, based on FFTs (figure 3 (a)), and the cochleagram or auditory spectrogram (figure 3(b)), derived from 128 “constant Q” cochlea filters. Note that the vertical axis on the auditory spectrogram plot indicates the filter number from 1 to 128

and not the actual frequency. Each filter number corresponds to a resonant frequency or characteristic frequency (CF) of the cochlea filter, and the CFs of the filters are arranged in a log frequency scale unlike the FFT spectrogram, which is linear. For this example, the CFs range from sub hertz to approximately half of the sampling rate, which correspond roughly to $[0, 500]$ Hz. The log frequency arrangement of the “constant Q” filters mimics the frequency response of the basilar membrane and inner hair cells of the cochlea. Further details can be found in papers by Shamma.^{3, 10, 11} The key difference between the two spectrograms is the fine temporal structure, or details of the signal, that is preserved in the auditory spectrogram. It is the fine temporal structure that provides the ITD information used by stereausis to extract DOA estimates.

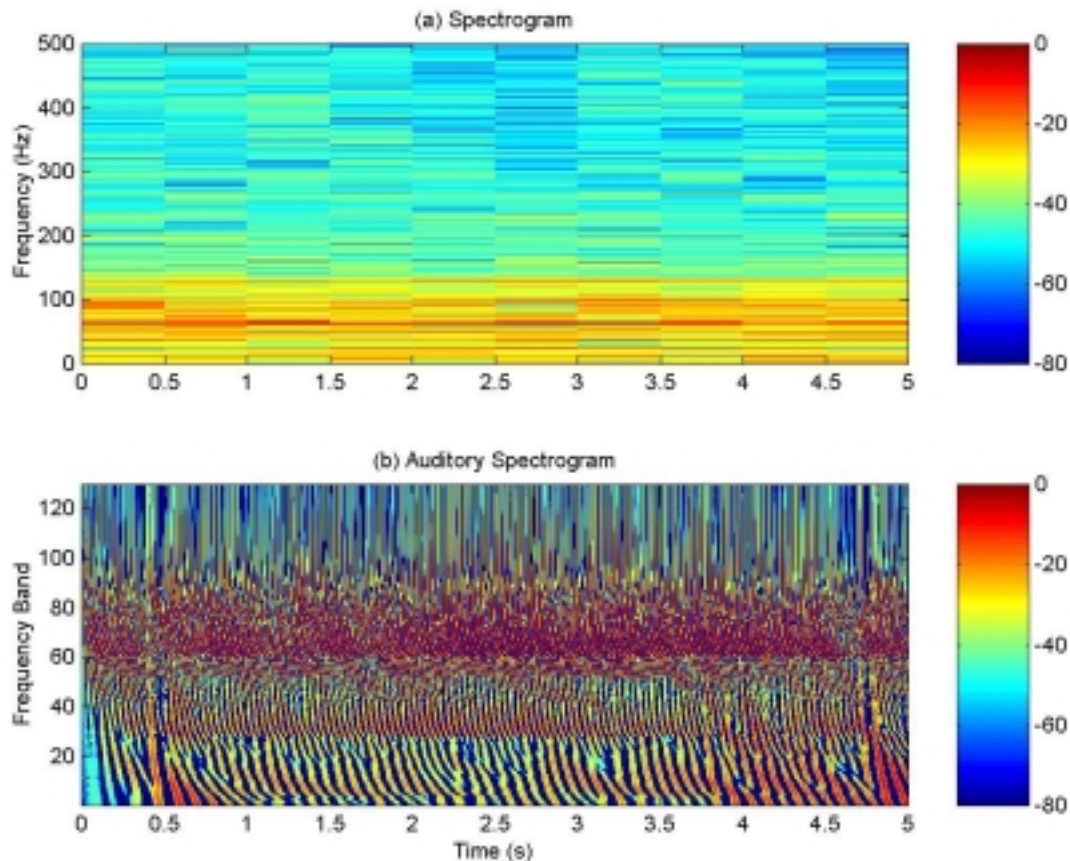


Figure 3. (a) Spectrogram and (b) cochleagram or auditory spectrogram of a moving tank.

3.1. GENERATING IMPOVERISHED SIGNALS

We have shown previously that using only three sensors from the seven-element array, we can obtain good DOA results for single and multiple vehicles using stereausis.^{5, 6} We have not fully exploited the fact that the human auditory system has a unique ability to process sound in a noisy environment (e.g., the cocktail party effect).

¹⁰ S. Shamma, “Speech processing in the auditory system I: Representation of speech sounds in the responses of the auditory nerve,” *J. Acoust. Soc. Am.*, Vol. 78, pp. 1612-1621, 1985.

¹¹ S. Shamma, “Speech processing in the auditory system II: Lateral inhibition and the processing of speech evoked activity in the auditory nerve,” *J. Acoust. Soc. Am.*, Vol. 78, pp. 1612-1621, 1985.

Therefore, we want to compare an auditory-inspired algorithm such as stereausis versus a statistical algorithm such as IWM, to see if there are any gains in using stereausis for processing impoverished signals. For performance analysis, we will only be using a very high SNR single-target data run with GPS ground truth and vary the power of the additive white Gaussian noise (AWGN). There are several ways to artificially inject noise into the signal. We chose to first calculate the average signal power for the entire run (approximately 150 s of data) and then use it as a reference level to generate various SNR cases. From second to second, the actual SNR level will be different. Figure 4 shows the spectrogram of the original signal and the -5 dB SNR case. For the -5 dB SNR example, the actual SNR will be a lot lower than -5 dB when the target is far from the sensor array, $[60, 150]$ s, and a lot higher than -5 dB when the target is near the closest point of approach (CPA), $[15, 20]$ s.

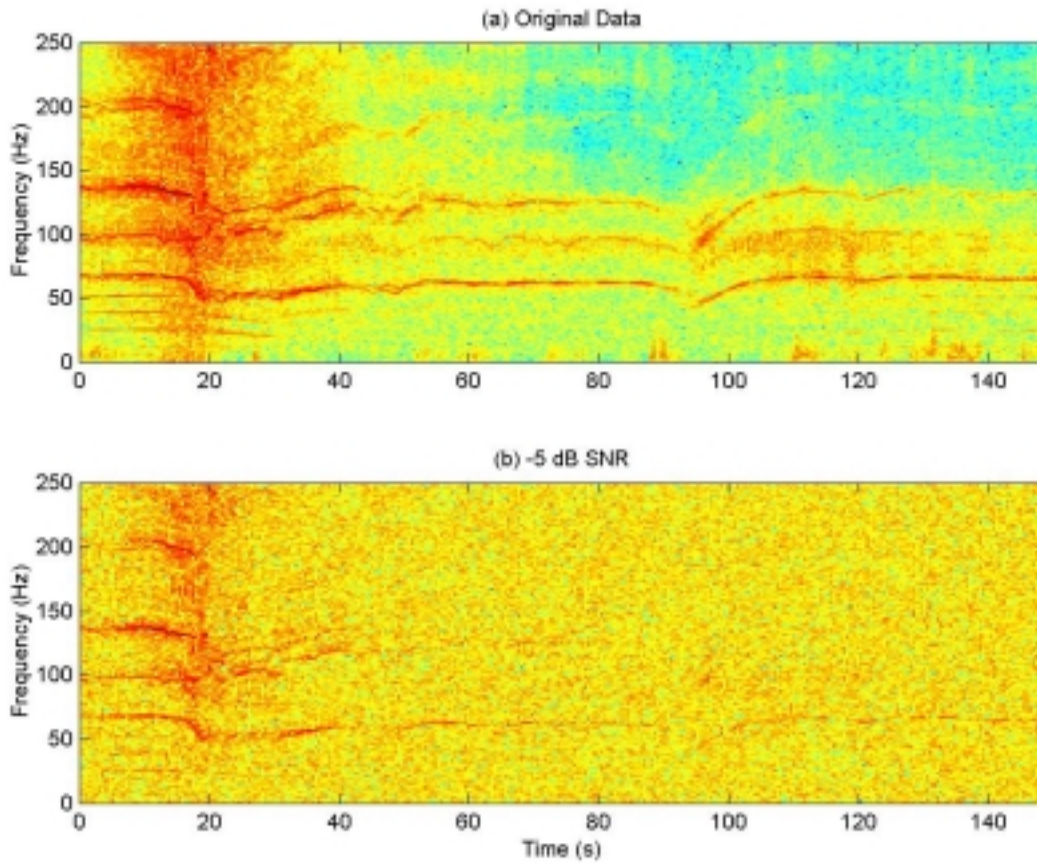


Figure 4. (a) Spectrogram of the original data and (b) spectrogram of -5 dB SNR case.

3.2. SENSOR ARRAY ISSUES

Direct one-to-one comparison between IWM and stereausis, however, is difficult because MUSIC requires many sensor array elements to accurately determine the signal and noise subspaces, while stereausis only requires a pair of sensors. We use three sensors (three pairs) for stereausis to help resolve the front-back ambiguity. Our analysis shows that using all possible pairs of microphones from the seven-element circular array does not seem to improve the performance of stereausis much. Therefore, we conduct the comparison using the seven-element circular array for IWM and three-element triangular array for stereausis as shown in figure 5.

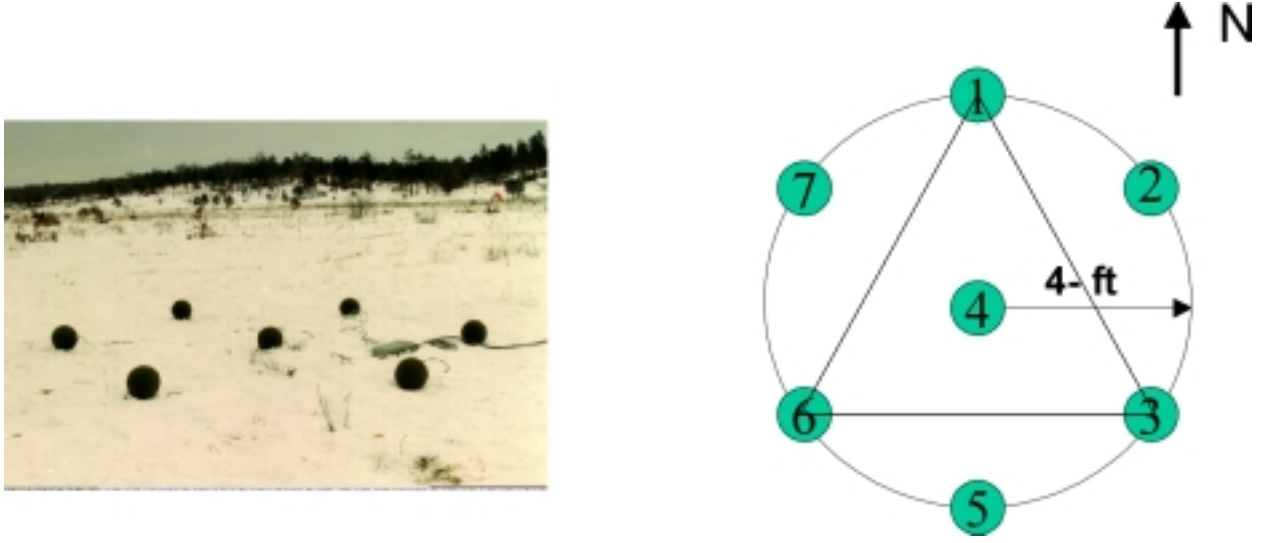


Figure 5. A seven-element circular array is used for IWM and a three-element triangular array (microphones number 1, 3 and 6) is used for stereausis.

3.3. ALGORITHM PERFORMANCE AND EXPERIMENTAL RESULTS

Performance results in terms of beampattern sharpness and DOA accuracy are discussed for four cases: original signal, 0 dB, -5 dB, and -10 dB. At and beyond -10 dB, both algorithms completely break down. For IWM, we use 20 largest frequency components from [20, 25] Hz for each 1-s frame of data, and the signal decomposition is adaptively determined at each frequency component. Note that we do not use the assumption that only one target can occupy a frequency component, as assumed previously.^{1, 2} For stereausis, the correlation measure used is $C(x_i, y_j) = x_i y_j$, and ITDs are determined by combining phase delays across peak frequency components in the stereausis plot (which is actually the disparity plot). The peak frequencies are extracted from the auditory power spectrum, obtained by collapsing the five diagonal lines to the left and five diagonal lines to the right of the main diagonal onto the main diagonal (see figure 2). Figures A1 to A8 in the appendix show beampattern results and the DOA estimates extracted from the maximum peak of the beampattern at each 1-s frame for ICM and stereausis for four SNR cases. We calculate the mean squared errors (MSE) of the DOA estimates with respect to the GPS ground truth for the 150-s data segment, whose spectrogram is shown in figure 4 (a). Table 1 shows the DOA MSE results with the corresponding number of outliers. We define an outlier as the DOA estimate at time k that satisfies the condition $\text{DOA_error}_k = |\text{DOA_est}_k - \text{GPS}_k| > 30$ degrees.

Algorithm	Original	0 dB	-5 dB	-10 dB
IWM	15.6 (10)	57.4 (29)	59.8 (65)	49.2 (105)
Stereausis	16.5 (6)	46.3 (31)	50.3 (85)	48.1 (110)

Table 1. DOA MSEs for IWM and stereausis for four SNR cases: original signal, 0 dB, -5 dB, and -10 dB. The number of outliers for each case is shown in parenthesis.

4. CONCLUSIONS

We have presented simulation and experimental results comparing a statistical approach, IWM, and an auditory-inspired approach, stereausis, for DOA estimation of ground vehicles. There are other, and perhaps better, ways for comparing the two approaches (e.g., using the exact same array), but we chose the current comparison method because we wanted to emphasize the strength of both algorithms. Based on the results in table 1, the performance, in terms of DOA accuracy, is comparable. Both algorithms performed poorly for the -5 dB and -10 dB cases, as indicated by the large number of outliers. Beampattern sharpness comparison is not easy, because each algorithm produces different types of patterns. IWM yields a more continuous beampattern, while stereausis yields a few peaks in the beampattern. Overall (see Figures A1-A8), both types of beampatterns degenerate dramatically at low SNR. In terms of real-time implementation, the main advantage of IWM over stereausis is lower computational complexity. However, if the aVLSI cochlea filters are readily available, computational complexity will not be an issue. The main advantage of stereausis over IWM is the number of sensors required. In some applications such as Small Unit Operation (SUO) robotic vehicle, it is a desirable system requirement to have only a few closely spaced sensors to perform localization.

Preliminary results show that auditory-inspired algorithms can be effectively applied to battlefield acoustic problems. However, there are still many fundamental issues to address to optimize and fully utilize these algorithms for applications other than speech.

APPENDIX

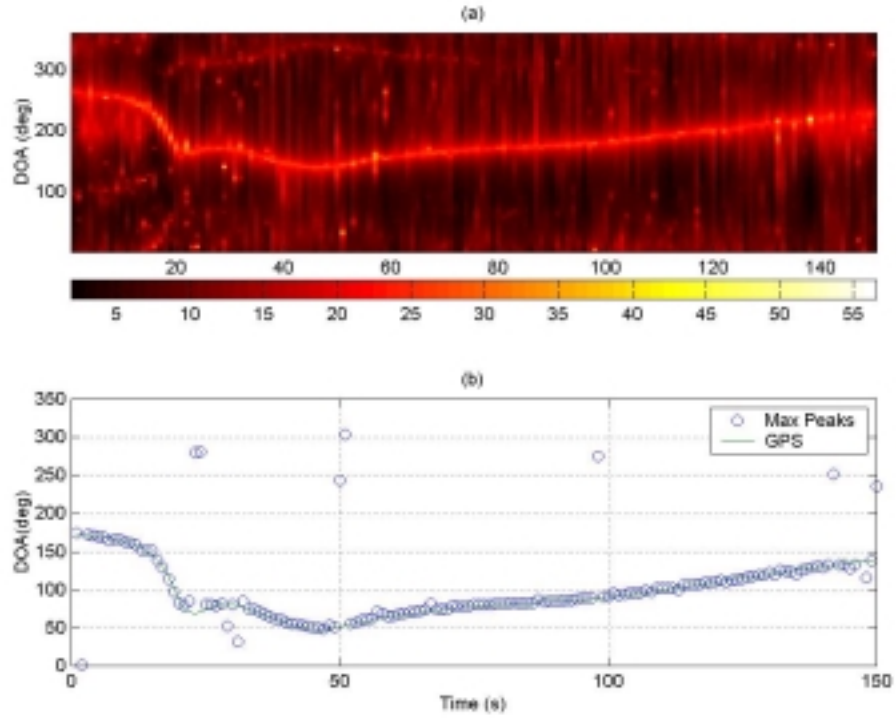


Figure A1. IWM applied to original data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

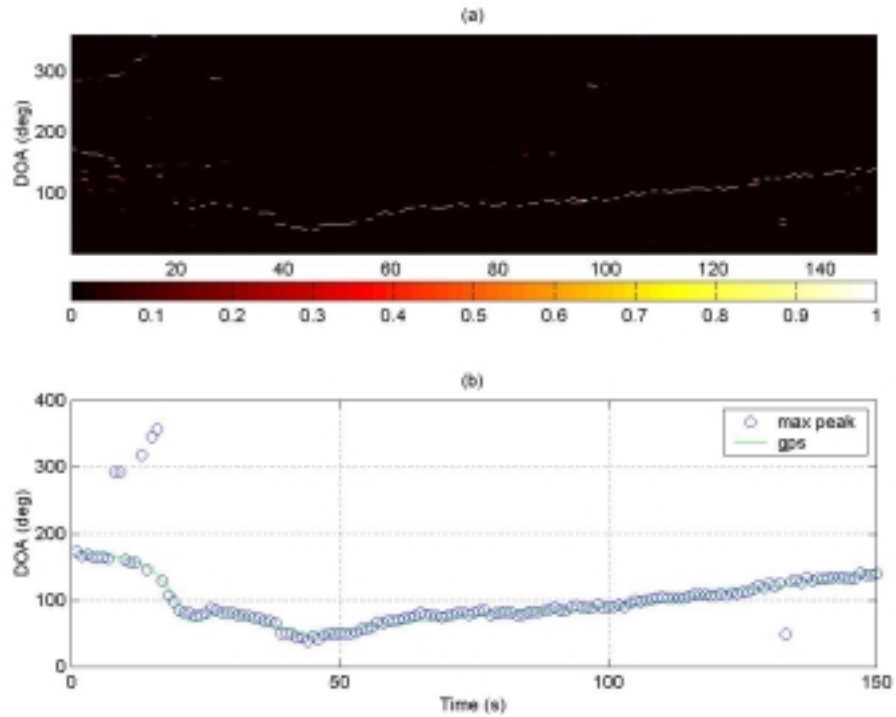


Figure A2. Stereausis applied to original data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

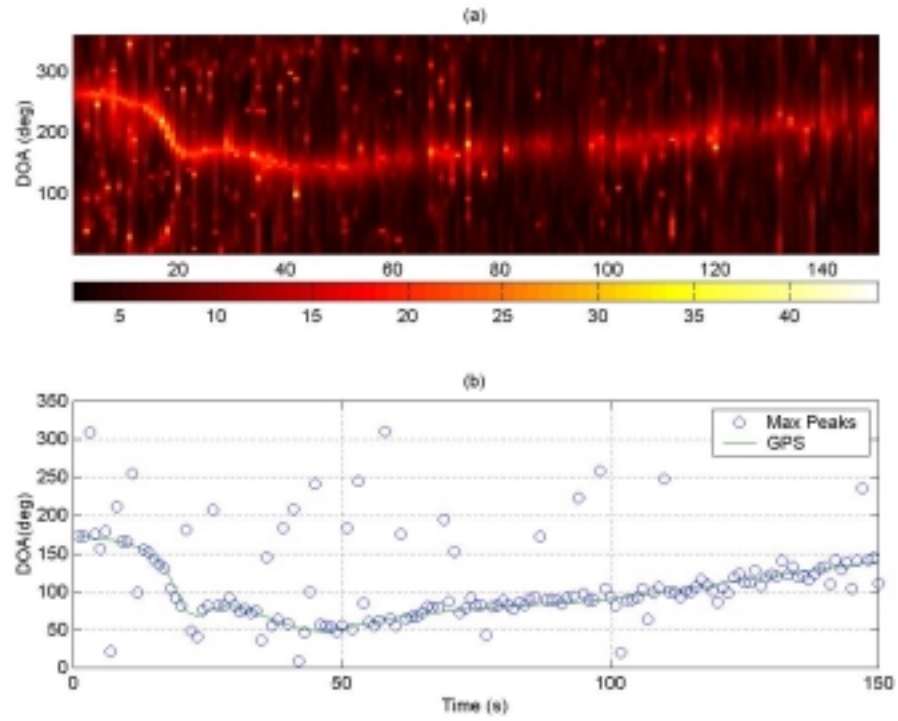


Figure A3. IWM applied to 0dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

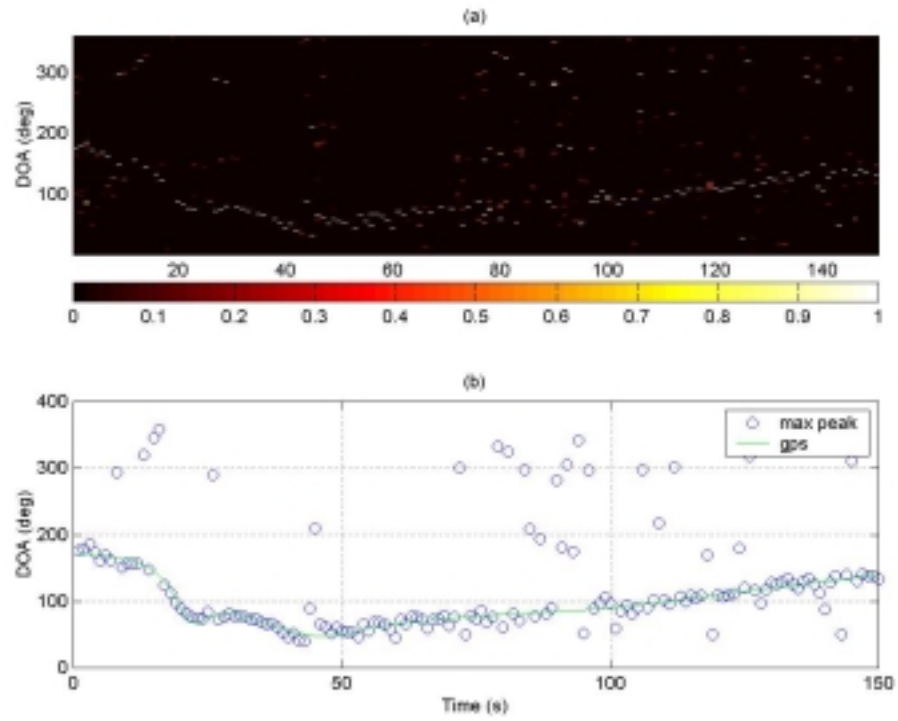


Figure A4. Stereausis applied to 0dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

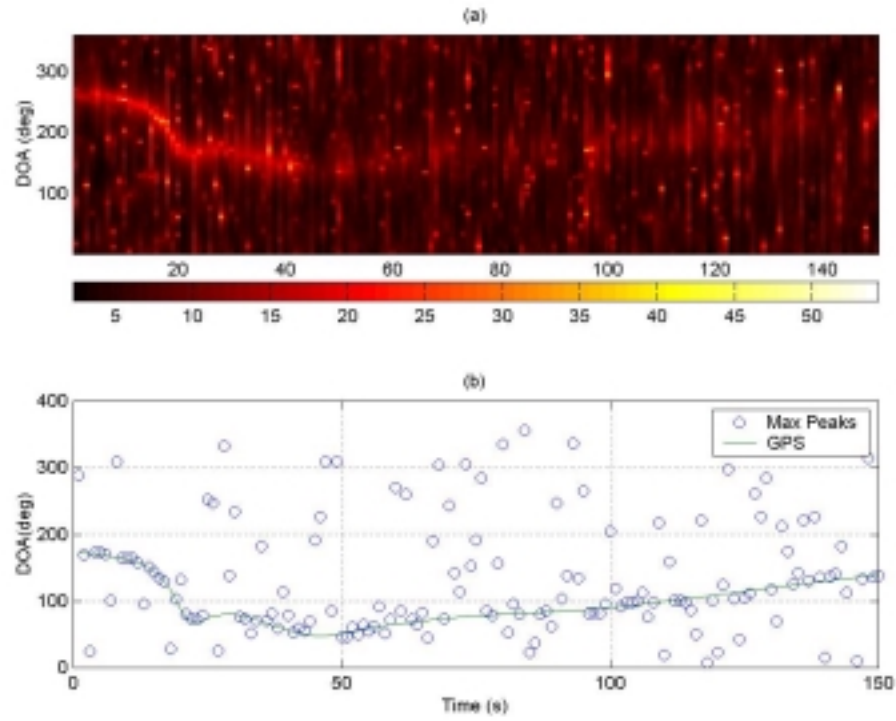


Figure A5. IWM applied to -5dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

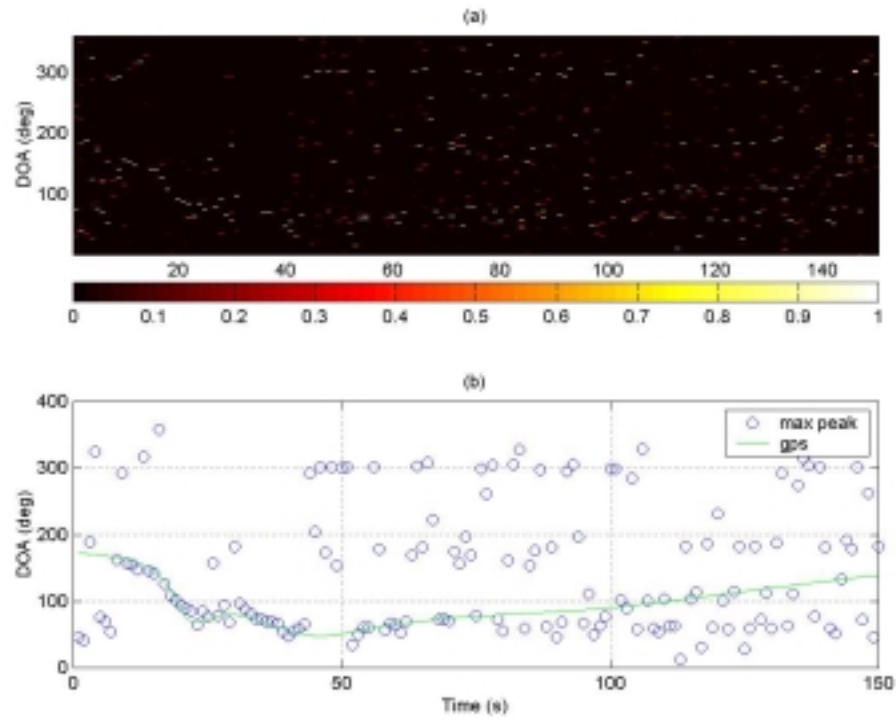


Figure A6. Stereausis applied to -5dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

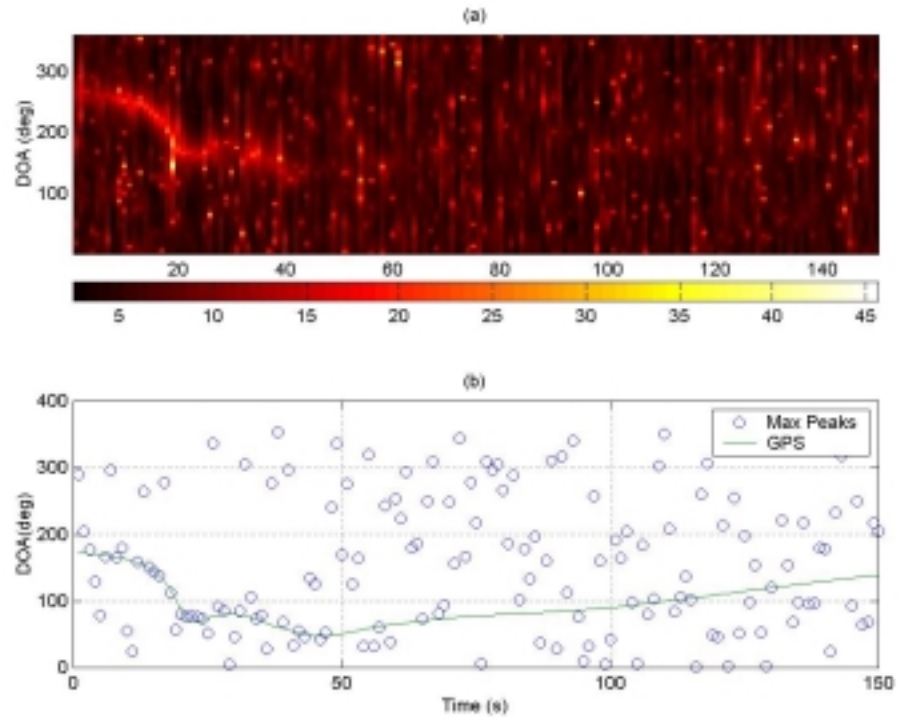


Figure A7. IWM applied to -10 dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.

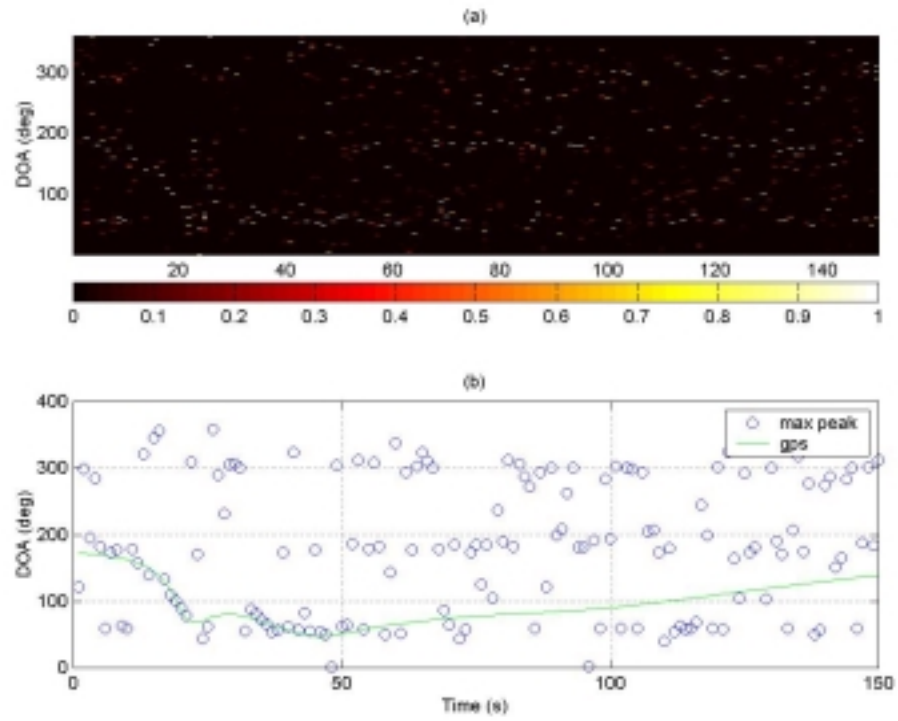


Figure A8. Stereausis applied to -10 dB SNR data: (a) Histogram of the beampatterns and (b) DOA estimates versus GPS.